Problem 1 (E-M - LIS exam '16)

1) Red / green / blue balls (n in total) drawn according to

$$P(Z_r = z_r, Z_g = z_g, Z_b = z_b) = C(z_r, z_g, z_b)\, 0.5^{z_r}\, \theta^{z_g} (0.5-\theta)^{z_b}.$$

C is constant w.r.t. θ. (Here, $C = C(z_r, z_g, z_b) = \frac{n!}{z_r!\, z_g!\, z_b!}$.)

We don't directly observe Z_r, Z_g, Z_b, but rather:

$$\begin{cases} \alpha = Z_r + Z_g \\ \beta = Z_b \end{cases}.$$

We want to use E-M to estimate θ. $(0 \le \theta \le 0.5)$

2) Let's assume that we had observed Z_r, Z_g, Z_b. Then, we can use max. likelihood to get an estimate for θ.

Log-likelihood: $\mathcal{L}(\theta) := z_r \log 0.5 + z_g \log \theta + z_b \log(0.5-\theta) + \log C$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{z_g}{\theta} - \frac{z_b}{0.5-\theta} := 0 \Rightarrow \boxed{\hat{\theta} = \frac{0.5\, z_g}{z_g + z_b}}$$

For example, $\left.\begin{array}{l} z_r = 1 \\ z_g = 2 \\ z_b = 1 \end{array}\right\} \Rightarrow \hat{\theta} = \frac{1}{3}$

(Did not use Lagrange multipliers for the constraint on θ, but need to make sure that $\hat{\theta}$ satisfies $0 \le \hat{\theta} \le 0.5$.)

3) E-M reminder: $\begin{cases} \text{Latent variables} & Z \\ \text{Observed variables} & X \end{cases}$

Ideally, want to $\max\limits_{\theta} \log P(x \mid \theta)$ [maximize the log-likelihood of the observations]

Since this is hard, we instead go for

$\max\limits_{\theta} \; Q^{(k)}(\theta)$ [maximize the expected log-likelihood of the observations]

$$\begin{cases} Q^{(k)}(\theta) := \mathbb{E}_{p^{(k)}}\left[\log P(x, z \mid \theta) \right] \\ p^{(k)}(z) := \mathbb{P}\left[z \mid x, \theta^{(k)} \right] \end{cases}$$

In our case, $z = (z_r, z_g, z_b)$

$X = (\alpha, \beta)$

$$\begin{cases} Q^{(k)}(\theta) = \mathbb{E}_{p^{(k)}}\left[\log P(\alpha, \beta, z_r, z_g, z_b \mid \theta) \right] \\ p^{(k)}(z) = \mathbb{P}\left[z_r, z_g, z_b \mid \alpha, \beta, \theta^{(k)} \right]. \end{cases}$$

Also, $\log P(\alpha, \beta, z_r, z_g, z_b \mid \theta) = \log P(\alpha, \beta \mid z_r, z_g, z_b, \theta) + \log(z_r, z_g, z_b \mid$

$= \underbrace{\log P(\alpha, \beta \mid z_r, z_g, z_b)}_{\text{does not depend on } \theta} + \log(z_r, z_g, z_b \mid$

Therefore, we will consider

$$Q^{(k)}(\theta) = \mathbb{E}_{p^{(k)}}\left[\log P(z_r, z_g, z_b \mid \theta) \right].$$

4) If $\xi_r^{(k)} := \mathbb{E}_{p(k)}[Z_r]$ and $\xi_g^{(k)} := \mathbb{E}_{p(k)}[Z_g]$, write $Q^{(k)}$ as a function of $\xi_r^{(k)}$, $\xi_g^{(k)}$, α, β, and θ.

$$Q^{(k)}(\theta) = \mathbb{E}_{p(k)}\left[\log C + Z_r \log 0.5 + Z_g \log\theta + Z_b \log(0.5-\theta)\right]$$

$$= \mathbb{E}_{p(k)}[\log C] + \xi_r^{(k)} \log 0.5 + \xi_g^{(k)} \log\theta + \beta \log(0.5-\theta).$$

5) $\boxed{\text{M-step}}$ Maximize $Q^{(k)}$ w.r.t. θ:

$$\frac{\partial Q^{(k)}}{\partial\theta} = \frac{\xi_g^{(k)}}{\theta} - \frac{\beta}{0.5-\theta} := 0 \Rightarrow \boxed{\theta^{(k+1)} = \frac{0.5\,\xi_g^{(k)}}{\xi_g^{(k)} + \beta}}$$

6) $\boxed{\text{E-step}}$ Compute $\xi_g^{(k)} = \mathbb{E}_{p(k)}[Z_g]$.

$Z_g \mid \alpha, \beta, \theta^{(k)} = Z_g \mid \alpha, \theta^{(k)}$ follows a binomial distribution with $n_{bin} = \alpha$ number of trials, and $p_{bin} = \dfrac{\theta^{(k)}}{0.5+\theta^{(k)}}$ probability of success.

Therefore, $\xi_g^{(k)} = n_{bin}\, p_{bin} \Rightarrow \boxed{\xi_g^{(k)} = \dfrac{\alpha\,\theta^{(k)}}{0.5+\theta^{(k)}}}$

Problem 2 (MC - HMM - LIS exam '15)

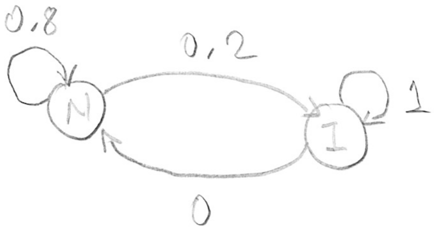
1) $Y_t \in \{I, N\}$

$$\begin{cases} P(Y_1 = N) = 1 \\ P(Y_1 = I) = 0 \end{cases}$$

$P(Y_{t+1} | Y_t)$

$Y_t \backslash Y_{t+1}$	N	I
N	0.8	0.2
I	0	1

Transition graph =



2) $P(Y_4 = N) = \sum_{y_1, y_2, y_3} P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, Y_4 = N)$

$$= \sum_{y_1, y_2, y_3} P(Y_4 = N | Y_3 = y_3) P(Y_3 = y_3 | Y_2 = y_2) P(Y_2 = y_2 | Y_1 = y_1) P(Y_1 = y_1)$$

all terms in the sum are 0, except for one

$$\underline{\underline{\quad}} \; P(Y_4 = N | Y_3 = N) P(Y_3 = N | Y_2 = N) P(Y_2 = N | Y_1 = N) P(Y_1 = N)$$

$$= 0.8 \times 0.8 \times 0.8 \times 1 = 0.512.$$

3) $P(Y_4, Y_6 | Y_5) \overset{?}{=} P(Y_4 | Y_5) P(Y_6 | Y_5)$.

$$P(Y_4, Y_6 | Y_5) = \frac{P(Y_4, Y_5, Y_6)}{P(Y_5)} \underset{\text{rule}}{\overset{\text{product}}{=}} \frac{P(Y_6 | Y_4, Y_5) P(Y_4, Y_5)}{P(Y_5)}$$

$$\underset{\text{property}}{\overset{\text{Markov}}{=}} \frac{P(Y_6 | Y_5) P(Y_4, Y_5)}{P(Y_5)} = P(Y_6 | Y_5) P(Y_4 | Y_5).$$

4) Hidden : $Y_t \in \{I, N\}$

Observed : $X_t \in \{W, D, F\}$

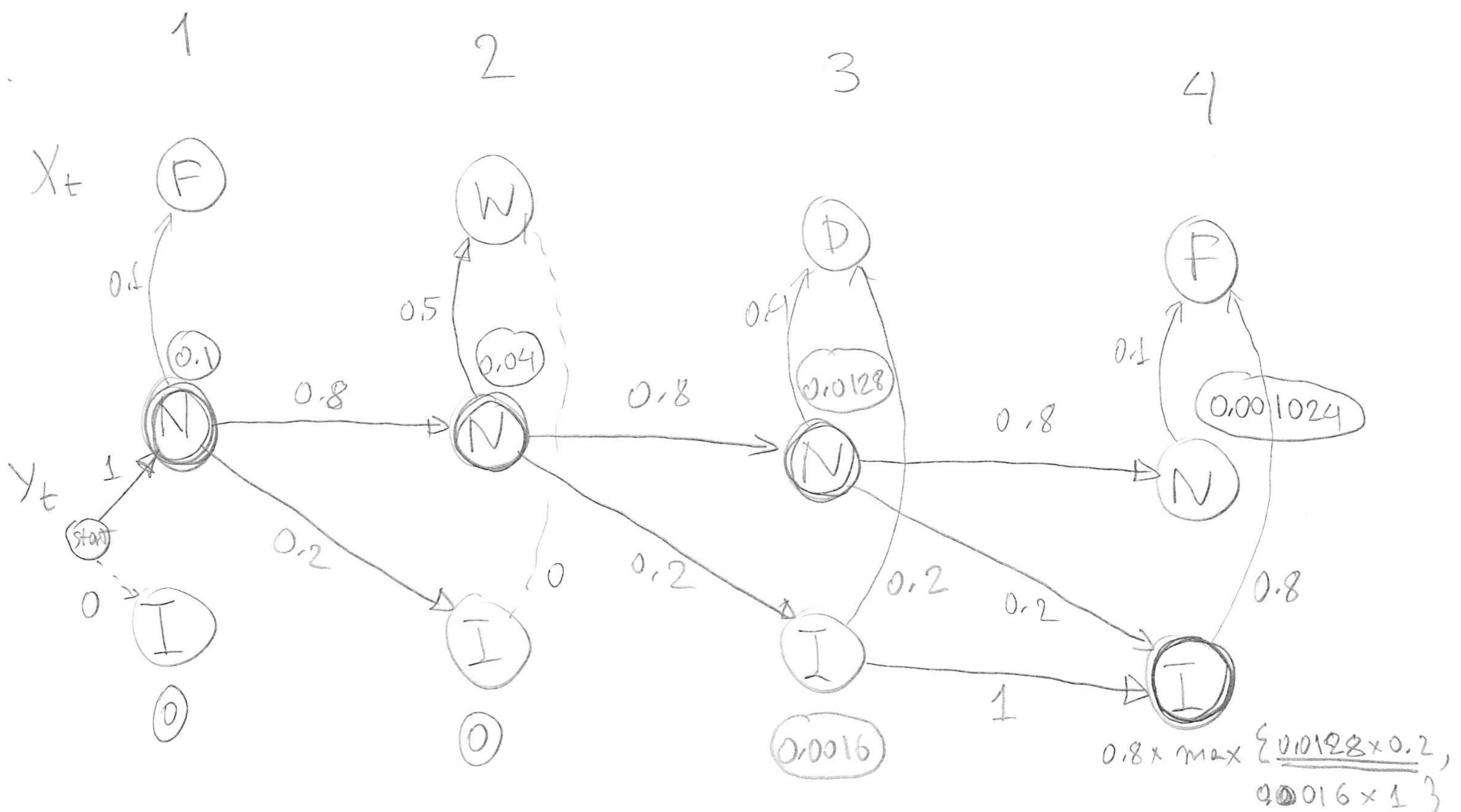
$Y_t \backslash X_t$	W	D	F
N	0.5	0.4	0.1
I	0	0.2	0.8

$P(X_t | Y_t)$

5) [Optional! This was not taught in the lectures and was not solved in the tutorial!]

Find $\underset{y_1, y_2, y_3, y_4}{\operatorname{argmax}} P(y_1, y_2, y_3, y_4 | X_1 = F, X_2 = W, X_3 = D, X_4 = F)$

Viterbi algorithm



$0.8 \times \max\{\underline{0.0128 \times 0.2}, 0.0016 \times 1\}$

$= 0.002048$

Most likely sequence: $\boxed{(\hat{y_1, y_2, y_3, y_4}) = (N, N, N, I)}$